DgEMM performance is data-dependent

Tom Cornebize, Arnaud Legrand
Inria Grenoble
09 September 2020, JLESC, the Internet
DGEMM PERFORMANCE IS DATA-DEPENDENT
...AND TIME-DEPENDENT

Tom Cornebize, Arnaud Legrand
Inria Grenoble
09 September 2020, JLESC, the Internet
Predictive simulation/emulation of High Performance Linpack (HPL).

https://hal.inria.fr/hal-02096571
Predictive simulation/emulation of High Performance Linpack (HPL).

https://hal.inria.fr/hal-02096571
Predictive simulation/emulation of High Performance Linpack (HPL).

Which one is wrong? Simulation model, platform calibration, or platform mis-configuration/unstability?

DGEMM takes 95% of HPL ⇒ need a lot of care for good predictions.

https://hal.inria.fr/hal-02096571
Predictive simulation/emulation of High Performance Linpack (HPL).

Which one is wrong? Simulation model, platform calibration, or platform mis-configuration/unstability?

DGEMM takes 95% of HPL ⇒ need a lot of care for good predictions.
Predictive simulation/emulation of High Performance Linpack (HPL).

https://hal.inria.fr/hal-02096571

Which one is wrong? Simulation model, platform calibration, or platform mis-configuration/unstability?

DGEMM takes 95% of HPL ⇒ need a lot of care for good predictions.
When and why does my platform change?
Performance non-regression tests

Evolution of the node dahu-26

- CPU 0
- CPU 1

Performance, but also frequency and temperature.

https://cornebize.pages.in2p3.fr/g5k_data_non_regression/
Performance non-regression tests

Evolution of the node dahu-14

- cpu: 0
  -冷却问题
  -冷却问题修复
  -BIOS升级

- cpu: 1
  -冷却问题
  -冷却问题修复
  -BIOS升级

图表展示了CPU在不同时间点的性能变化，包括冷却问题和BIOS升级的影响。
Performance non-regression tests

Overview of the cluster dahu

Probability
- [+] 0% - 0.01%
- [+] 0.01% - 0.1%
- [+] 0.1% - 1.0%
- [+] 1.0% - 10.0%
- 10.0% - 100%
- [-] 1.0% - 10.0%
- [-] 0.1% - 1.0%
- [-] 0.01% - 0.1%
- [-] 0% - 0.01%
Performance non-regression tests

Performance, but also frequency and temperature.

https://cornebize.pages.in2p3.fr/g5k_data_non_regression/
RANDOMIZING REALLY MATTERS!
 SHOULD WE RANDOMIZE THE CONTENT OF THE MATRIX?

DGEMM durations on the CPU 0 of dahu-1 (matrices of size 2048×2048)
Should we randomize the content of the matrix?

Evolution of DGEMM durations on the CPU 0 of dahu-1 (matrices of size 2048×2048)
Should we randomize the content of the matrix?
SHOULD WE RANDOMIZE THE CONTENT OF THE MATRIX?

Evolution of DGEMM durations on the CPU 0 of dahu-1 (matrices of size 2048×2048)

Evolution of the frequency on the CPU 0 of dahu-1
Should we randomize the content of the matrix?

Possible crazy explanation: bit flips consume more energy.
Testing the hypothesis: applying a mask to the random values

https://hal.inria.fr/hal-02401760

Images adapted from https://en.wikipedia.org/wiki/Double-precision_floating-point_format
Testing the hypothesis: applying a mask to the random values

Images adapted from https://en.wikipedia.org/wiki/Double-precision_floating-point_format
Testing the hypothesis: applying a mask to the random values

Distribution of DGEMM durations on the CPU 0 of dahu-1 (matrices of size 2048\times2048)

https://hal.inria.fr/hal-02401760

Images adapted from https://en.wikipedia.org/wiki/Double-precision_floating-point_format
When the computations mess with the communications
Interference between communications and computations

Durations of a 256 MiB MPI_Recv

- Locality
- Loopback
- Remote

Expected duration on a 100 Gbps link

Duration (seconds)

Idle

Computations on the node
Interference between communications and computations

Durations of a 256 MiB MPI_RECV

Expected duration on a 100 Gbps link

Locality
loopback
remote

Duration (seconds)

Idle
DGEMM
Computations on the node
Interference between communications and computations

Durations of a 256 MiB MPI_Recv

Locality
- loopback
- remote

Expected duration on a 100 Gbps link

Computations on the node
- Idle
- DGEMM
- MPI_Iprobe
INTERFERENCE BETWEEN COMMUNICATIONS AND COMPUTATIONS

Durations of a 256 MiB MPI_Recv

Duration (seconds)

Expected duration on a 100 Gbps link

Computations on the node

Idle DGEMM MPI_Iprobe MPI_Iprobe & DGEMM

Locality loopback remote
### Predictive simulation/emulation of High Performance Linpack (HPL)

Which one is wrong? Simulation model, platform calibration, or platform mis-configuration/unstability?

DGEMM takes 95% of HPL ⇒ need a lot of care for good predictions.

### Performance non-regression tests

Performance, but also frequency and temperature.

### Should we randomize the content of the matrix?

Possible crazy explanation: bit flips consume more energy.

### Testing the hypothesis: applying a mask to the random values

### Interference between communications and computations

Contact: tom.cornebize@inria.fr

https://hal.inria.fr/hal-02096571
https://cornebize.pages.in2p3.fr/g5k_data_non_regression/
https://hal.inria.fr/hal-02401760